SILVA BALCANICA

Silva Balcanica 26(2): 61–83 (2025)
doi: 10.3897/silvabalcanica.26.e159961

RESEARCH ARTICLE

Tree species distribution modelling using stand-level forest inventory data – a *Pinus sylvestris* L. case study in Bulgaria

Petar Dimitrov

Space Research and Technology Institute – Bulgarian Academy of Sciences, Acad. G. Bonchev St., Bl. 1, 1113 Sofia, Bulgaria

Corresponding author: Petar Dimitrov (petar.dimitrov@space.bas.bg)

Academic editor: Georgi Georgiev | Received 2 June 2025 | Accepted 11 August 2025 | Published 17 September 2025

Citation: Dimitrov P. 2025. Tree species distribution modelling using stand-level forest inventory data – a *Pinus sylvestris* L. case study in Bulgaria. Silva Balcanica 26(2): 61–83. https://doi.org/10.3897/silvabalcanica.26.e159961

Abstract

National Forest Inventory (NFI) data provided at the field plot-level represent a valuable data source for Species Distribution Modelling (SDM) as demonstrated by many studies. However, a stand-level forest inventory, conducted in some countries, including Bulgaria, is less appropriate because matching these data to spatial predictor datasets is not straightforward. This study aimed to propose and evaluate a methodology for tree species distribution modelling at a national scale driven by stand-level forest inventory data, focusing on *Pinus sylvestris* L. as a target species. The forest polygons were filtered using a set of attribute and geometric criteria and transformed to points to generate a presence/absence dataset. Based on the literature, four predictor variables were selected, including growing degree-days above 5° C (DD.5), August precipitation (PPT08), potential yearly global radiation (YRA), and soil volumetric coarse fragment content (CF60), all available as 250 m resolution rasters. Generalised Linear Models with different combinations of linear, quadratic, and interaction terms were tested. The effect of spatial errors in the presence/absence dataset was investigated by randomly shifting the points up to 1 km from their original position. Evaluation statistics were calculated by 5-fold spatial block cross-validation repeated 10 times. The basic model, which included only linear terms, performed the worst. The other models were more or less equal performers. The model, which included a quadratic term for DD.5, was therefore selected as the optimal model as it had the most parsimonious structure. The Area under the Receiver Operating Characteristic curve and the True Skill Statistic for that model were 0.80±0.04 and 0.49±0.08, respectively. The model was relatively robust to spatial errors in the occurrence data as simulated by randomly shifting the points. The modelled distribution was much wider than the actual natural distribution of the species, and all mountain areas were

Copyright *Petar Dimitrov*. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

identified as potential distribution areas based on these four predictors. As drawbacks of the approach should be noted the assumption of homogeneity of forest stands, the unknown reliability of the inventory dataset, and the low prevalence. Nevertheless, the study demonstrated successfully the utility of the stand-level NFI dataset for SDM.

Keywords
Cross-validation, GIS, GLM, mapping, raster, response curve, Scots pine, SDM

Introduction

The current geographic distribution of tree species and the tree composition of forests are the result of a complex combination of factors, including environmental conditions, biotic interactions, and past disturbance events (Radkov, 1961). Species distribution patterns are of interest to ecologists and forest managers alike, seeking answers to theoretical and practical problems such as how different environmental variables control the species occurrence or abundance, or under which environmental conditions a successful cultivation of a species is possible (Falk, Mellert, 2011). However, the intrinsic complexity of the processes involved makes species distribution patterns hard to explain and predict. Species Distribution Modelling (SDM) is used increasingly both to improve our understanding of factors shaping species distribution and to map areas of *potential* present or future distribution. SDM relates species observations to environmental variables using statistical or machine learning methods (Guisan, Thuiller, 2005). When environmental spatial data layers are available, the modelled species-environment relationships (response surfaces) can be used for predictive (distribution) mapping (Franklin, 2009) over an area of interest. Depending on their theoretical grounds, different authors may use alternative terminology to describe such models and their output, including habitat suitability (Guisan, Thuiller, 2005; Benito Garzón et al., 2006) and ecological niche modelling (ENM; e.g. Pshegusov et al., 2022). Elith and Leathwick (2009) suggest using the more neutral term SDM rather than ENM. SDM have been widely used for mapping current and future distribution of trees at continental, regional, and national scales (Benito Garzón et al., 2006; Armenteras, Mulligan, 2010; Zimmermann et al., 2013; Prasad et al., 2020; Pshegusov et al., 2022; Bonannella et al., 2022; Mauri et al., 2022).

Pinus sylvestris L. (Scots pine) has a vast areal that extends across Eurasia (Rehfeldt et al., 2002; Houston Durrant et al., 2016). In Bulgaria, *Pinus sylvestris* forests have a larger area than any other conifer-dominated forest type, and most of them are situated in the Rhodope, Rila, Pirin, and Plana Mountains, with smaller areas also present in several other mountains (Bondev, 1991). However, it is mostly absent from the Stara Planina Mountains. Most of the natural forests are situated between 1000 m and 1900 m a.s.l. However, the species can be found naturally in lower elevations (about 700 m a.s.l) as well as up to 2200 m a.s.l (Panayotov,

Yurukov, 2016). Artificial *Pinus sylvestris* stands are also present, many of them at altitudes lower than the natural lower limit of the species in Bulgaria (Panayotov, Yurukov, 2016). The species is usually described as being tolerant to a variety of conditions, including temperature extremes, high and low levels of soil moisture and fertility (Radkov, 1961). However, Panayotov et al. (2023) note that, at the tree line, the combination of low temperatures, strong winds, and snow accumulation causes damage, and under such extreme conditions, *Pinus peuce* and *Picea abies* are more competitive. It requires high light levels and would not withstand shadowing (Radkov, 1961). In mountains, it occupies mostly southerly exposed slopes (Panayotov, Yurukov, 2016), rarely other expositions, for instance, when the soil is poor enough to prevent the growth of concurrent species (Radkov, 1961). Scots pine forests predominantly occupy the upper part of slopes, which are drier, whereas the wetter zones, at lower places, with rich soils are occupied by *Fagus sylvatica* (Gogushev, Zlatanov, 2016). Anthropogenic activity, such as primitive gold and iron metallurgy from antiquity to XIX century, the associated charcoal production, and the clearing of forests by fire to be used as pasture have contributed to the distribution of Scots pine by providing open space for it to colonise (Radkov, 1961). Such secondary forests may occur even in northerly expositions, for example in the Rila Mountains (Radkov, 1961).

It was shown that *Pinus sylvestris* seed viability correlated significantly with the mean June–August temperature at the tree line in Sweden (Kullman, 2007). Therefore, the low regeneration success at high altitude, where the summer season is short and cold, may be one of the factors limiting the species‘ distribution. Another unfavourable factor, which restricts the altitudinal range limit and high-elevation abundance of pine trees, is winter desiccation (Kullman, 2007), a phenomenon observed also in dwarf pine (*Pinus mugo*) in Rila Mountain (Bencheva, Doychev, 2022). At low elevations, the increased risk of summer drought stress probably influences the distribution. It was shown that Scots pine from Rila Mountain responded with radial growth reduction to a dry climate period (Dimitrov et al., 2011). Bigler et al. (2006) showed that multi-year drought reduces tree growth for several years or decades, thus increasing a tree's risk of death. It is assumed that the widespread occurrence of high Scots pine mortality in some central alpine dry valleys of the European Alps is related to such drought events.

In many countries, National Forest Inventory (NFI) datasets are available, which include data collected in field plots distributed according to a regular grid or other sampling design (Vidal et al., 2016). These datasets are a typical source of species observation data for tree distribution modelling because plots' spatial co-ordinates are known and because both presence and absence of a species are registered in each plot. Forest management and inventory in Bulgaria are carried out following a regional principle. The forest territory is distributed among several Regional Forest Directorates, and each of them is subdivided into smaller territorial units, "*Darjavni Gorski Stopanstva*" and "*Darjavni Lovni Stopanstva*" (DGSs/DLSs). Forest inventory is performed in each DGS/DLS periodically, usually at a

10-year interval. Inventories of separate DGSs/DLSs are not synchronous. In addition, forest inventory is performed in national parks and other protected areas following similar procedures. An output of the inventory is a digital dataset comprising polygons and their descriptions. The polygons represent the boundaries of forest stands, forest roads and other types of land currently not occupied by trees but part of the forest fund (collectively referred to as "*podotdeli*"). For the forest stands, the descriptive part includes parameters like height, diameter, and age for each tree species in each forest layer, growing stock volume, soil type and structure, slope exposition, etc. In some cases, plot-level data are collected as part of the inventory field work, but final results are reported at stand-level (Markof, Karamfilov, 2016). The data are not so suitable for SDM because instead of a "point" observation, they represent an area within which the species has been observed. Therefore, the assumption has to be made that the species is distributed in all parts of the area, which is plausible because, by definition, each forest stand should be homogeneous in terms of forest characteristics to allow effective management. Despite this complication, stand-level forest inventory is a valuable data source with potential application in SDM. A recent study by Panayotov et al. (2023) used such data to model the ecological niches of the main tree species in South-western Bulgaria under current and future climate. In their study, the spatial unit of modelling and predictive mapping was the forest stand. With this approach, the mapped area is the area covered by the forestry data (areas outside existing forests are not considered), and the spatial resolution is relatively coarse and not constant as stands' area vary. An advantage is that soil and other environmental data from the inventory can be used because they are available for every mapping unit. Another strategy, followed in the present study, would be to use a Geographic Information System (GIS) tool to find the centroid or a representative point of each polygon (forest stand) and use these points as a simulated plot-level dataset. This would permit predictive mapping over the territory of Bulgaria with constant spatial resolution, depending on the explanatory variables data used.

Therefore, this study aimed to propose and evaluate a methodology for tree species distribution modelling at a national scale driven by stand-level forest inventory data, focusing on *Pinus sylvestris* L. as a target species. Moreover, this methodology should permit mapping at medium spatial resolution (250 m), suitable for modelling species assemblages and understanding species occurrence patterns on a medium scale.

Datasets

Environmental variables

The selection of explanatory environmental variables was based on theoretical considerations and prior literature, rather than automated statistical procedures. This approach was used in previous studies (e.g. Franklin, 1998). The four cho-

sen variables were growing degree-days above 5° C (DD.5), August precipitation (PPT08), potential yearly global radiation (YRA), and soil volumetric coarse fragment content (CF60). Growing degree-days were selected because this variable describes the thermal conditions, which are among the most important factors for tree growth (Bolliger et al., 2000). The northern limits of many woody species correlate with total annual growing degree days (Sykes et al., 1996). GDD is frequently used in SDM of trees (Huntley et al., 1995; Bolliger et al., 2000; Zimmermann et al., 2007; Guisan et al., 2007; Meier et al., 2010; Marchi, Ducci, 2018; Prasad et al., 2020). August precipitation describes water availability and serves as an indicator of potential summer drought intensity. Other studies have also employed variables characterising water availability during the summer period to predict tree distribution. For instance, Zimmermann and Kienast (1999) used a moisture index (the difference between precipitation and potential evapotranspiration) for July, while Bolliger et al. (2000) utilised a water budget for the same month. Panayotov et al. (2023) and Benito Garzón et al. (2006) demonstrated that summer precipitation is a crucial factor for modelling Scots pine distribution. We opted for August precipitation since it was the driest month throughout most of the study area. We did not use a complex aridity index which compares temperature and precipitation because GDD already accounts for the thermal conditions and because we study its interaction with PPT08 during modelling. Potential yearly global radiation indicates the energy input, which largely determines potential productivity (Leathwick, Austin, 2001). The content of coarse fragments in the soil was used as a proxy of several other soil characteristics. Higher coarse fragment content is typically associated with poorly developed (or eroded), thin soils at higher altitudes and steeper slope angles. Coarse fragments affect soil physical properties such as water retention and porosity (Brakensiek, Rawls, 1994; Chow et al., 2007; Baetens et al., 2009). Soil fertility generally decreases with increasing sand and coarse fragment content, while aeration improves (Radkov, 1961). Soil stoniness has been reported as one of the important factors for modelling Scots pine distribution (Panayotov et al., 2023). It is also used as one of the criteria for assessing soil richness and humidity in field mapping of forest site types (Executive Forest Agency, 2011).

The PPT08 and DD.5 were calculated using the ClimateEU v 4.63 software (Marchi et al., 2020). Provided with geographic coordinates and elevation data, ClimateEU interpolates a set of climatic variables for any point (or a set of points) in Europe. The spatial resolution is not fixed or restricted, and we set it to 250 m, as in Marchi and Ducci (2018), taking into account a balance between fine scale and plausible interpolation results. A square grid of this size was created, with the centroids of the grid cells used as input points for ClimateEU. The elevation of the cells (and the respective points) was calculated as the average of the 30 m Copernicus Digital Surface Model (DSM) dataset (Copernicus, 2024). The interpolated climate data were transformed into raster format with a pixel size of 250 m. They represented the climate conditions for the period 1961–1990 to which

Pinus sylvestris populations are putatively adapted. Daily potential (i.e. clear-sky) solar radiation was calculated using the r.sun module of GRASS GIS (Hofierka, Suri, 2002) for each 30 m pixel of the Copernicus DSM (Copernicus, 2024). The calculated values represent the global radiation, i.e. the sum of direct, diffuse, and reflected radiation, and take into account topographic shadowing. To reduce the overall computation time needed to obtain yearly estimates, we ran r.sun for a single day each month, then multiplied by the number of days as an estimate of monthly radiation, and finally summed the monthly values to estimate yearly solar radiation (Franklin, 1998). The final raster was resampled to 250 m by taking the average. Coarse fragment data at a spatial resolution of 250 m were obtained from the SoilGrids 2.0 dataset (Poggio et al., 2021) for four soil depth ranges (layers): 0-5 cm, 5-15 cm, 15-30 cm, and 30-60 cm. They represent the volumetric content of fragments larger than 2 mm in units of cm^3/dm^3. To obtain a single coarse fragment value for the layer 0-60 cm, the weighted average of the four layers was calculated, with the layer‘s relative thickness serving as the weighting factor; for instance, the weight for the 30-60 cm layer is 0.5 as it represents half of the total depth of 60 cm.

Presence/absence dataset

Presence and absence data for *Pinus sylvestris* were obtained by processing 173 forest inventory datasets. Most of the data (170 datasets) represented inventories of DGSs/DLSs. The other included the territories of the national parks “Rila” and “Centralen Balkan”, and the “Rilomanastirska gora” nature reserve. The data for the “Pirin” national park were not used because the descriptive part was available in pdf format only. Data were missing or corrupted also for some DGSs/DLSs. Nevertheless, when combined, the datasets covered most of the Bulgarian territory and the areal of the species. We used the most recent inventory for a particular territory that was available and could be successfully processed. The processing included transformation of the data from the native plain-text *.zem format to the ESRI Shapefile format. The processing of the geometries (i.e. polygons) and the descriptive data was performed using custom PYTHON 3.12 scripts (Python Software Foundation, https://www.python.org/). The shapefiles were projected from the Bulgarian Coordinate system 1970 to the Coordinate system BGS2005 using BGSTrans v4.6.2. (GCCA, 2020) and then to WGS84 / UTM 35N using QGIS. Data was then filtered using several attributes and geometric criteria. The aim was to retain only the stands in which vegetation was close to the natural condition (in terms of tree species composition) and which had a compact shape and relatively small area. First, the “*Type of podotdel*” attribute was used to select stands of type “*Seed stand*”, “*Coppice stand*”, and “*Pinus mugo stand*”. Therefore, artificial stands and other types of polygons, such as forest roads, were discarded. Second, a stand was discarded when an invasive species was present in it (one of *Robinia pseudoacacia*, *Ailanthus altissima*, *Gledicia triacantus*, *Acer negundo*, *Laburnum anagyroides*,

and American ash) and also when any species was marked as having an artificial origin. Third, a polygon was retained if its area was no greater than 0.0625 km^2, the elongation was below 0.7, and the convexity was above 0.6. Convexity is the ratio of the polygon's area to the area of its convex hull. Elongation (E) is calculated by: E = 1 - S / L, where S and L are, respectively, the lengths of the short-axis and the long-axis of the minimum rotated rectangle enclosing the polygon. The maximal area threshold (i.e. 0.0625 km^2) was determined to correspond to the spatial resolution of the predictors. The rationale for using the geometry criteria was to remove very large (and potentially inhomogeneous) stands as well as stands with narrow, long, and complex form. With this, we are trying to minimise the uncertainty associated with using a single point to represent the polygon. A list of tree species was made for each stand passing the above filtering criteria. The list included only species present in the first or the second forest layer (that is, young trees, seedlings, saplings, etc., are ignored) and explicitly marked as of natural origin (that is, ignoring species with no data about origin). Finally, a representative point was created in each polygon using the corresponding function in the PYTHON's GEOPANDAS package (Bossche et al., 2023). The representative point is conceptually similar to the centroid but has the advantage that it is guaranteed to be within the polygon. At this stage, the 173 point datasets were merged into a single national dataset. Points with *Pinus sylvestris* were marked as presence and all the other - as absence. We additionally created artificial absence points in the high mountain areas above the tree line. These were generated in a regular 500 m grid in the areas above 2200 m a.s.l. assuming that above that altitude, there should be no trees. This altitude was selected based on the maximal altitude (2185 m a.s.l.) of the representative points (except those with *Pinus mugo*).

After this initial selection, the point density was still high. It was not uncommon for more than one point to fall within a single 250 m 250 m cell of the predictors' rasters. To avoid the redundancy without the need to decide on which point to retain if more than one is present in the cell, we continued the analysis at the cell level. That is, the raster cells' centroids were computed and marked as presence if the corresponding cell contained even one presence point and as absence if it contained only absence points (the centroids of cells without points were discarded). Using the centroids of the raster cells instead of the representative points does not introduce further uncertainty of geographic location because the SDM is performed at the cell level anyway. The presence/absence dataset thus formed (n=265,036) is shown in Figure 1.

To study the effect of the maximum area filtering criteria, an alternative dataset was prepared using the same procedure but with a maximum area set to 0.25 km^2, therefore allowing larger stands (n=330,364). In addition, several modifications of this dataset were prepared to simulate different levels of error in the coordinates of the points. The points were randomly shifted in one of the four cardinal directions at distances of 250, 500, 750, and 1000 m. The random shift was done 10 times for each distance.

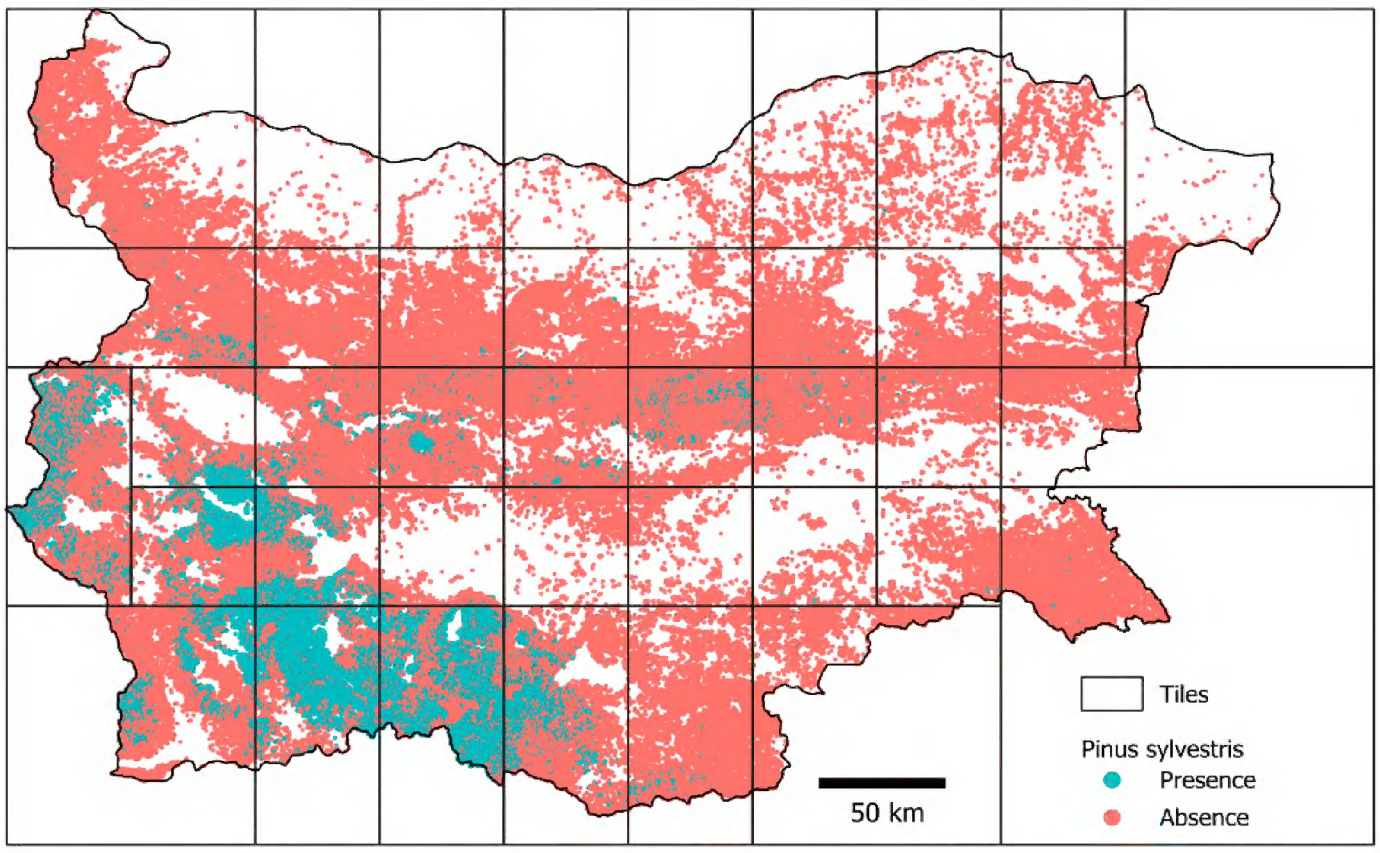

Fig. 1. Presence/absence dataset for Scots pine in Bulgaria (n=265,036). The study area was divided into 40 tiles used for cross-validation of the model

Methods

Subsampling

Subsampling is used with some species occurrence datasets to deal with autocorrelation (Guisan, Thuiller, 2005) and sampling bias (Varela et al., 2014). Our dataset has low prevalence (frequency of presences) and points are spatially clustered (high density in mountain areas and low density in lowlands). To mitigate these problems, the data was subsampled. Different subsampling strategies have been suggested in the literature, including environment (Varela et al., 2014) and geographic filtering methods (Bonannella et al., 2022). Here, geographic filtering was applied by overlaying the territory with a 5 km by 5 km grid and selecting one point at random in each grid cell. The procedure was applied to presence and absence points separately.

Descriptive statistical analyses

The subsampled dataset was examined for correlation between the environmental variables by means of Spearman's correlation coefficient. Histograms were used to visually evaluate the similarity or dissimilarity of the variables' distributions between the presence and absence data points. They also showed the range of the vari-

ables. Histograms that show the proportion of presence in each histogram bin were also analysed. They show, under particular value of the environment variable, what is the proportion of points that have the species and can be used to suggest the form of the relationship between each environmental variable and the response variable (Hosmer, Lemeshow, 2000).

Modelling

We used Generalised Linear Models (GLMs, Nelder, Wedderburn, 1972) to predict the distribution of *Pinus sylvestris*. This method has been widely used for SDM (Guisan et al., 2002). For a particular GLM, one should define: (i) a probability distribution of the response (an error function), (ii) a linear predictor, and (iii) a link function (Austin et al., 1990). Based on the nature of the response variable, an appropriate probability distribution should be selected. This should be a distribution that belongs to a family of distributions called the exponential dispersion model family (Dunn, Smyth, 2018), with examples including the normal, binomial, Poisson and gamma distributions, among others. It is assumed that each outcome Y of the response is generated from the selected distribution. The linear predictor and the link function together describe how the explanatory variables are related to the mean of the response μ (Dunn, Smyth, 2018). The linear predictor η has the following general form:

$$\eta = \beta_0 + \Sigma_{j=1}^{p} \beta_j x_j,$$

where β_0 is a constant term, β_j (for $j = 1, 2, \ldots p$) are p regression parameters, and x_j (for $j = 1, 2, \ldots p$) are p explanatory variables. The link function, $g()$, is a function linking η to the mean of the response μ so that $g(\mu) = \eta$.

In the case of SDM of presence and absence data, when the response has only two possible values (0 and 1), it is appropriate to assume that it has a binomial distribution (Nicholls, 1989). Accepting the binomial distribution, what we are modelling is the probability of an observation being 1, i.e. the probability of presence. Hence, the modelled responses have values changing continuously between 0 and 1. The link function usually used with the binomial distribution for SDM is the logit function:

$$\ln [\mu/(1-\mu)]=\eta,$$

where ln is the natural logarithm (Austin et al., 1990). The link maps the real number line, $(-\infty, +\infty)$, to the interval (0, 1). Using the logit link with a first-order polynomial linear predictor will give a S-shaped response curve (Austin et al., 1990). Higher order polynomial terms, or other parametric transformations, can be included in the linear predictor to account for more complex forms such as bell-shaped and multi-modal responses (Guisan et al., 2002; Elith, Leathwick, 2009).

Six models with different structures of the linear predictor were compared (Table 1). They all included the four explanatory variables, but with different combinations of linear, quadratic, and interaction terms. The model "m0" is the basic model where all terms are linear. The models "q1" and "q2" additionally include quadratic terms for DD.5 and both DD.5 and CF60, respectively. This allows for bell-shaped response curves for the two variables, a form which was suggested based on a preliminary examination of the data. The models "i1", "i2", and "i3" include interaction terms between PPT08 and the other three variables. It is known that the effect of precipitation on soil moisture differs with temperature, soil texture, exposition, and slope (Radkov, 1961). For instance, at the same precipitation, soil moisture is higher under lower temperatures (Radkov, 1961). In our case, it is hypothesised that the effect of PPT08 depends on DD.5, CF60, and YRA.

Table 1. Description of the alternative GLMs tested in this study

Model identifier	Description	The formula used in the R function *glm()*
m0	Linear terms	BialBor ~ DD.5 + PPT08 + CF60 + YRA
q1	Quadratic term for DD.5	BialBor ~ DD.5 + I(DD.5^2) + PPT08 + CF60 + YRA
q2	Quadratic terms for both DD.5 and CF60	BialBor ~ DD.5 + I(DD.5^2) + PPT08 + CF60 +I(CF60^2) + YRA
i1	Quadratic term for DD.5 and interaction between DD.5 and PPT08	BialBor ~ (DD.5 + I(DD.5^2)) * PPT08 + YRA + CF60
i2	Quadratic terms for both DD.5 and CF60 and interaction between CF60 and PPT08	BialBor ~ DD.5 + I(DD.5^2) + PPT08*(CF60 +I(CF60^2)) + YRA
i3	Quadratic term for DD.5 and interaction between YRA and PPT08	BialBor ~ DD.5 + I(DD.5^2) + PPT08*YRA + CF60

Note: Abbreviations of the explanatory variables: DD.5 - growing degree-days above 5° C, PPT08 - August precipitation, YRA - potential yearly global radiation, CF60 - soil volumetric coarse fragment content

Evaluation

We used several statistics to evaluate the models. These included the threshold-independent statistics Area under the Receiver Operating Characteristic (ROC) curve (AUC) and correlation coefficient, and the threshold-dependent statistics kappa, true negative rate (TNR, also called specificity), true positive rate (TPR, also called sensitivity), and true skill statistic (TSS, where TSS = sensitivity + specificity - 1).

The threshold-dependent statistics were calculated based on the threshold that maximised the sum of specificity and sensitivity (Jiménez-Valverde, Lobo, 2007). The reader is referred to Allouche et al. (2006) for details and formulas of these statistics.

We applied 5-fold spatial block cross-validation repeated 10 times to obtain out-of-sample evaluation statistics for the six candidate models from Table 1. The study area was divided into tiles with a size of 50 km by 50 km. Tiles at the border of the country were united with the adjacent within-country tiles so that no tile has too few points, resulting in the final set of 40 tiles shown in Figure 1. The tiles were assigned to 5 folds at random. In each iteration, one fold (8 tiles) was used for validation, and the remaining 4 folds (32 tiles) were used for training. The assignment of tiles to folds was repeated 10 times. Geographic filtering was applied in each iteration. In each iteration, the threshold value derived from the training data was used for calculating evaluation statistics from the validation data. Boxplots were used to compare the evaluation statistics for the candidate models. In addition, differences among models were evaluated by ANOVA and post-hoc tests (Tukey Honest Significant Differences (HSD)) applied to evaluation statistics values. When distribution was not normal according to the Shapiro-Wilk test, the nonparametric Kruskal-Wallis test (Kruskal, Wallis, 1952) was used to compare the models. In addition, a null model approach similar to that proposed by Bohl et al. (2019) was used to evaluate the results. The data points of the initial dataset (i.e. before subsampling, Figure 1) were randomly assigned to absence (0) or presence (1). The data were then subsampled as described above and used to calibrate GLM. The predictions of the model were compared with the true absence/presence label of each point and evaluation statistics calculated. The procedure was repeated 100 times and mean and standard deviation of the evaluation statistics were reported.

A final model was prepared using the full dataset (all tiles) and the best linear predictor structure and used to create a raster with the predicted probabilities of presence and a binary presence/absence raster over the study area. Data analysis was carried out in R (R Core Team, 2021) using the build-in *glm()* function to parameterise the GLMs, the TERRA package (Hijmans, 2023) to handle the vector and raster geospatial data and make the predicted probabilities raster, and the *pa_evaluate()* function from the PREDICTS package (Hijmans, 2024) to calculate the evaluation statistics.

Results

The application of geographic filtering led to a substantial reduction of the number of points (from 265,036 to 5359) and an increase in the prevalence (from 0.18 to 0.25). Using this subsampled dataset, the correlations between predictor variables were examined. Highest correlations were observed between DD.5 and CF60 (-0.69) and between DD.5 and PPT08 (-0.59). The correlation between CF60 and PPT08 was 0.34. The YRA was practically not correlated with the other variables.

With increasing altitude from the lowlands to the mountains, the DD.5 decreases, whereas the CF60 and PPT08 tend to increase. The histograms of the environmental variables (Figure 2, top) show overlap between the presence and absence groups. The modes of the two groups roughly coincide for PPT08 and YRA, while for DD.5 and CF60, they differ slightly. The lower four graphs in Figure 2 show the proportions of presence in each histogram bin. For DD.5 and CF60, the proportions are zero at the two ends of the gradient, from where they increase towards the middle, reaching a more or less uniform plateau. This form can be roughly approximated by a bell-shaped curve. Concerning the other two variables, the proportions of presence show a more complex, bimodal pattern. Interestingly, at the right extreme of YRA, all or most points are presence points.

The distributions of evaluation statistics derived from the cross-validation are shown in Figure 3. The "m0" model had the lowest median values of the evaluation statistics, except for TPR. Though, the boxplots do not suggest superiority of any of the other models. The ANOVA results showed that there was no significant difference between the models except for kappa ($p<0.05$) and correlation ($p<0.001$). Kruskal-Wallis test was carried out for TPR, also indicating the lack of significant difference between models. The Tukey HSD showed a difference between kappa for "m0" and "i3" ($p<0.01$), and between correlation for "m0" and the other five models ($p<0.001$). The base model "m0" performs the worst because it does not account for curvilinear relationships. However, none of the other mod-

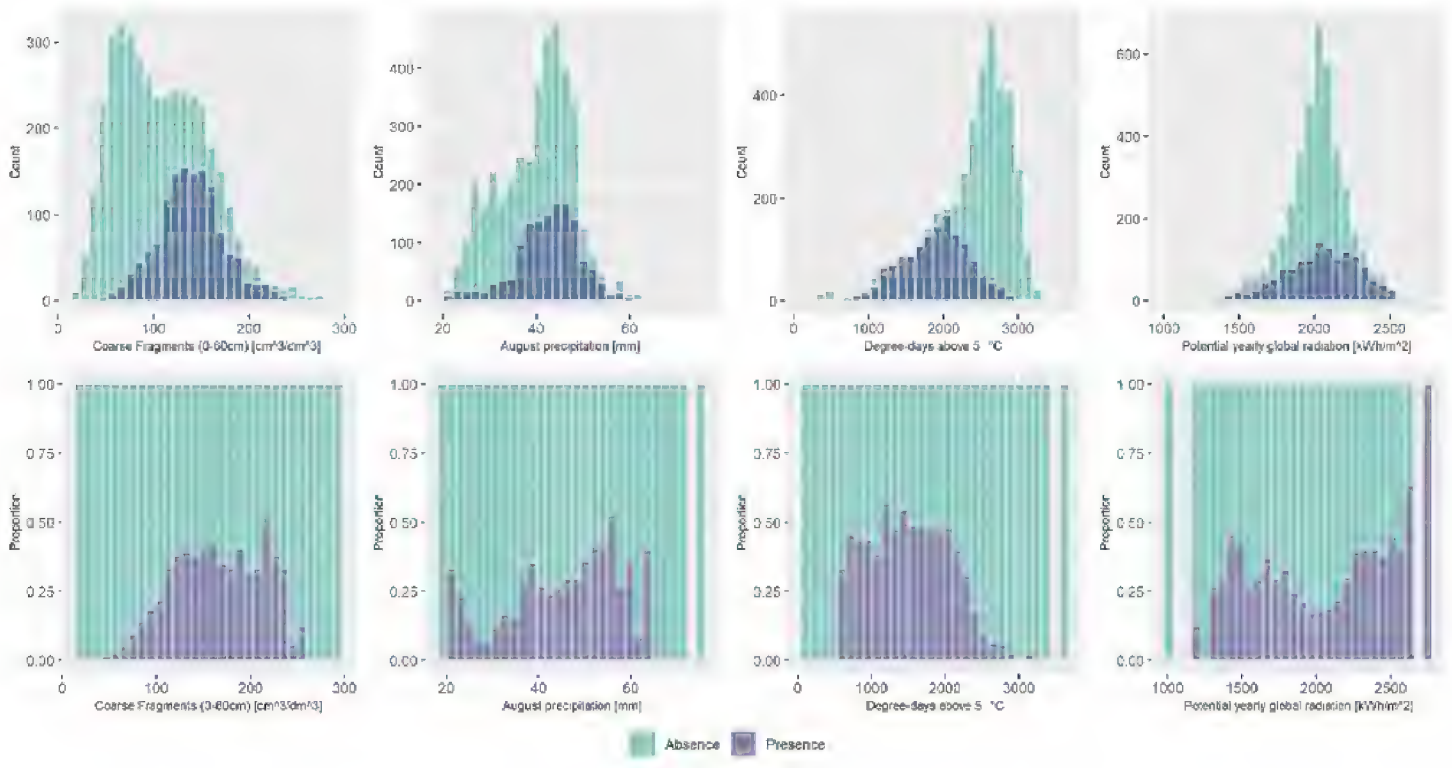

Fig. 2. Histograms of the environmental variables showing counts (above) and proportions for every bin (below) of presence and absence points (data after geographic filtering; n=5359, prevalence=0.25)

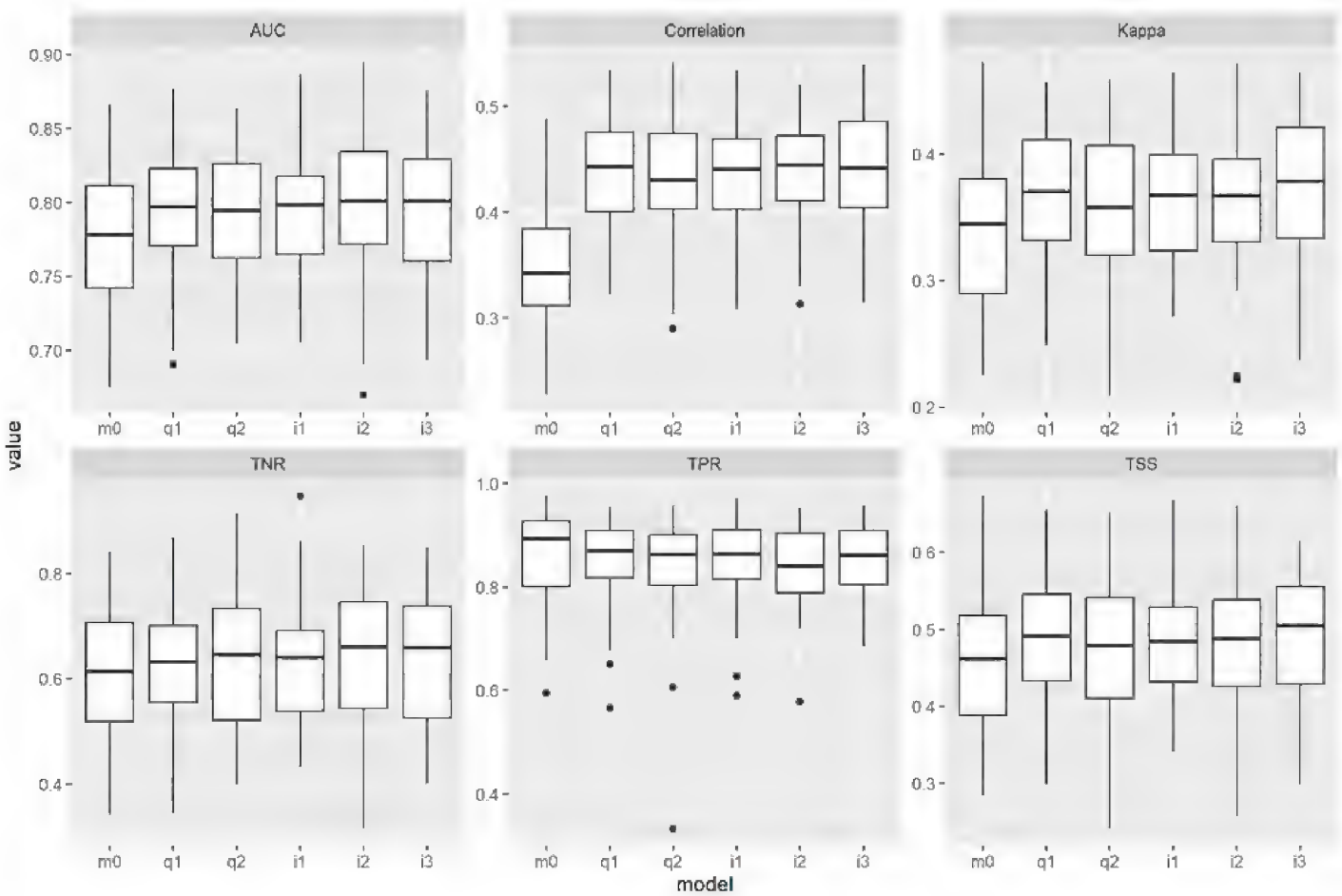

Fig. 3. Distributions of the six evaluation statistics (individual plots) for the models described in Table 1. The data used represent the result from 5-fold cross-validation repeated 10 times (n=50). AUC: Area under the receiver operating characteristic (ROC) curve; TNR: true negative rate; TPR: true positive rate; TSS: true skill statistic

els stand out. Therefore, we select "q1" as the optimal model form as it has the most parsimonious structure (fewer number of parameters). The AUC and TSS mean (standard deviation) for that model were 0.80 (0.04) and 0.49 (0.08), respectively. The same statistics for the null model were 0.49 (0.18) and 0.14 (0.16). A final model of the form "q1" parameterised using data from all tiles was used to derive a raster layer with the predicted probability of presence and a binary presence/absence map (Figure 4).

The final model was compared with the model parameterised with the alternative dataset, which used the 0.25 km^2 maximum area filter. The in-sample AUC and TSS for the 0.25 km^2 dataset are 0.798 and 0.503, whereas for the 0.0625 km^2 dataset, they are 0.804 and 0.521. There is practically no difference in model performance between the two datasets. This can be explained by the fact that only 20% of the points in the alternative dataset are new compared with the 0.0625 km^2 dataset. These points represent forest stands with an area between 0.0625 km^2 and 0.25 km^2. The random shifting of points led to a decrease in AUC and TSS (Figure 5). AUC seems to be more sensitive to even a small amount of shifting. However, even with the points shifted by 1000 m, the accuracy was not much deteriorated.

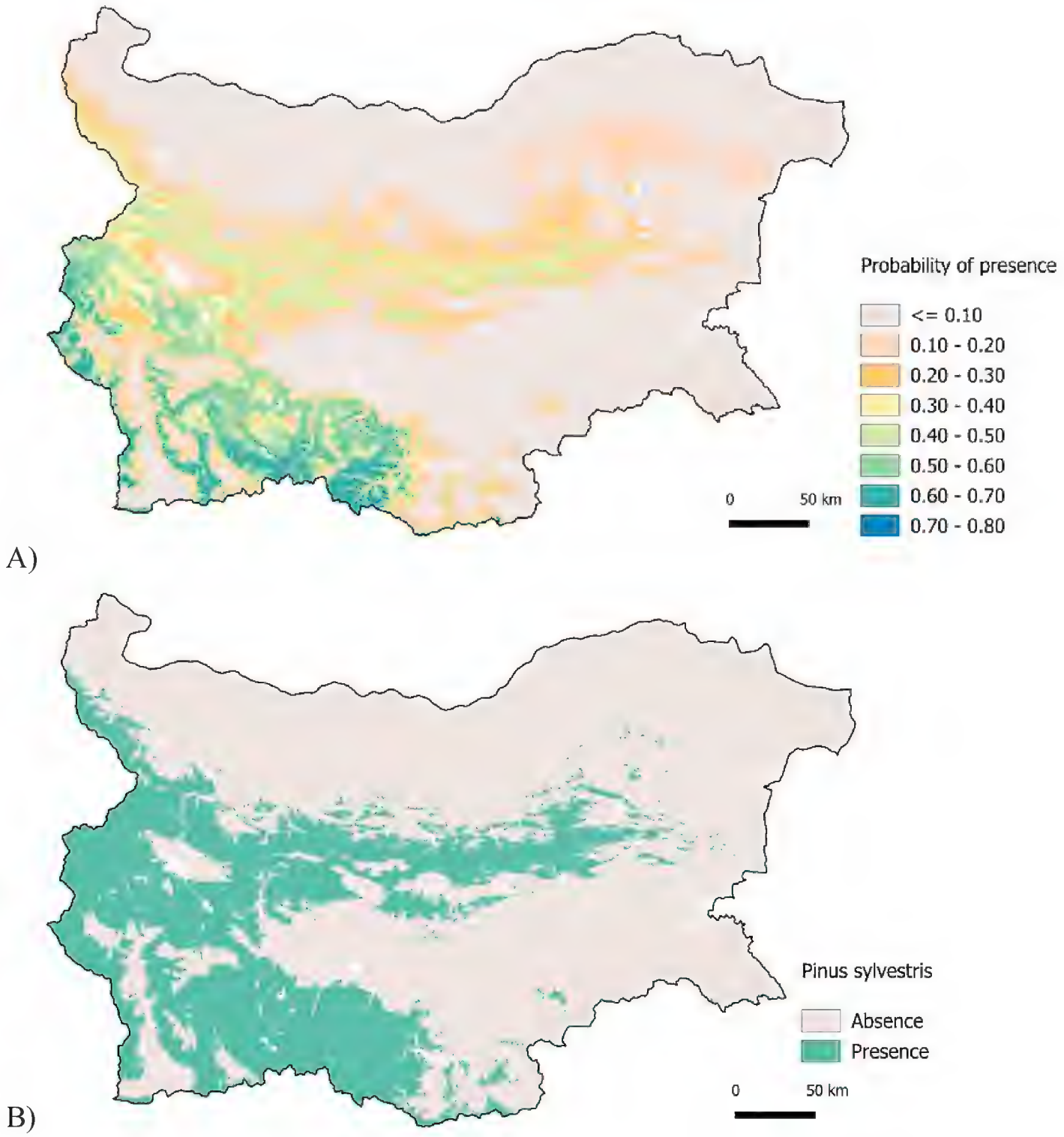

Fig. 4. A) Predicted probability of presence for *Pinus sylvestris* in Bulgaria based on the final GLM (the final model is model "q1", Table 1, with data from all tiles used for parameterisation). B) Binary map after applying a 0.216 threshold (the threshold that maximised the sum of specificity and sensitivity) on the probabilities in A

Discussion

This study tested the utility of stand-level forest inventory data for tree species distribution modelling at 250 m resolution for Bulgaria, focusing on *Pinus sylvestris* as an example. The forest polygons of the original dataset were filtered using a set of attribute and geometric criteria and transformed to points to generate a presence/absence dataset for modelling. GLM with four explanatory variables, selected a pri-

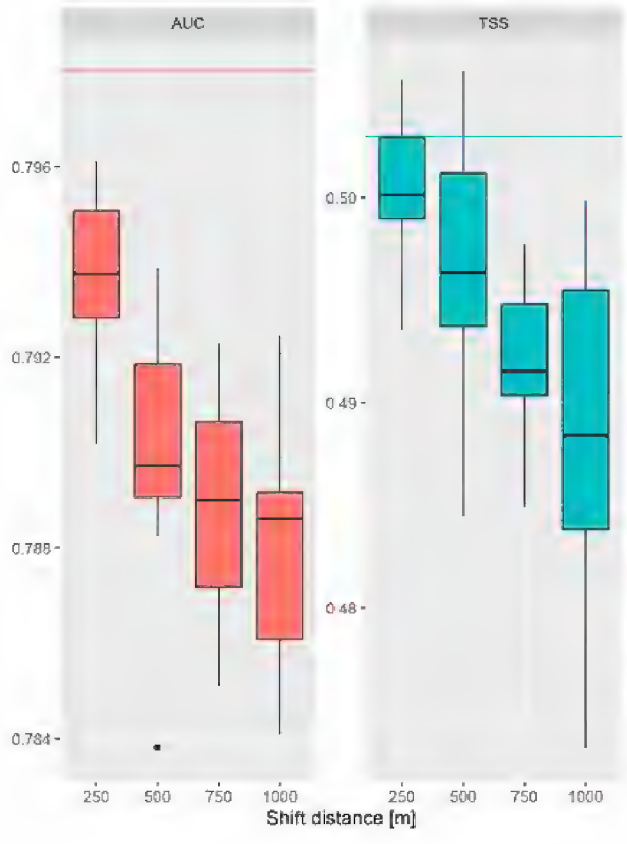

Fig. 5. Distribution of the in-sample AUC and TSS across 10 repetitions for each shifting distance. The horizontal lines indicate the in-sample AUC and TSS for the model without point shifting (0.798 and 0.503, respectively). The model used is q1 (see Table 1); the dataset used is this with a 0.25 km^2 maximum area filter

ori and available as spatial datasets, was used to establish the potential distribution areas of *Pinus sylvestris*. The modelling accuracy was good, reaching, for the optimal model, mean AUC=0.80 and mean TSS=0.49 according to the cross-validation results. According to the binary map (Figure 4B), the modelled distribution was much wider than the actual natural distribution of the species. All mountain areas were identified as potential distribution areas. Nevertheless, the probability of presence map (Figure 4A) shows variation within that area, with consistently lower probabilities in the Stara Planina and Sredna Gora ranges. Northern expositions also tend to have lower values for the predicted probability of presence. The overestimation of the area of presence probably indicates a significant effect of factors not accounted for in the model, such as biotic interactions and disturbances. Modelling the distribution of *Pinus sylvestris* in the Iberian Peninsula, Benito Garzón et al. (2006) obtained a similar overestimation result. They argue that the larger predicted area may be due to the inclusion of areas that are suitable for the pine and where it was present in the past. Their hypothesis was supported by the presence of palynological data showing the recent presence of Scots pine in some mountain ranges, where it is now practically extinct. While an area may be suitable, it may not be occupied. The actual distribution of a species could be the result of different factors, including dispersal and colonisation, migration rates of species, habitat fragmentation and historic factors (Benito Garzón et al., 2006).

The appropriate extent and grain of SDM is dictated by the study goals, the system, and available data (Elith, Leathwick, 2009). In this study, one of the goals was to make predictions that would permit modelling species assemblages and understanding medium-scale species occurrence patterns, such as those observed at different expositions and slope positions in mountain areas or along river courses. For this aim, the spatial resolution (grain) of 1 km, typically used in continental and regional-scale studies, is not appropriate. However, the available climatic datasets usually have 1 km resolution and are difficult to use with variables derived from DEM, such as YRA, which permit much finer grain. Interpolation of climate databases with ClimateEU allows for finding a balance. The 250 m grid cell gives a realistic interpolation result and was successfully applied in a previous study (Marchi, Ducci, 2018). Ideally, the resolution of the species ground data and the predictors should be the same or comparable (Guisan, Thuiller, 2005). To ensure spatial matching between the predictors and the species occurrence, we used only forest polygons with an area of 0.0625 km^2, which corresponds to the area of a 250 m by 250 m square. In addition to the matching spatial scale, exact colocation of the ground and predictor datasets is also important, particularly for sessile organisms, such as trees, and for studies performed at finer resolution (Guisan, Thuiller, 2005). However, Graham et al. (2008) found that useful predictions of species distributions can be made even when occurrence data include some spatial error. Marchi and Ducci (2018) achieved good results with plot data for which exact coordinates were not available, but only the 1 km grid cell in which the plot was located. Similarly, our results with points shifted by up to 1 km showed slight and gradual deterioration in accuracy and therefore some robustness to spatial errors in occurrence data. This supports the conclusion that the transformation of stand-level inventory datasets to point datasets through a simple GIS operation is a practical and justified approach to generate occurrence data for SDM.

The scale not only depends on the goal and the available data but also poses requirements for the type of predictors used. This is because different environmental factors operate at different spatial and temporal scales. As a consequence, their relative importance in defining a species' distribution and abundance can be highly scale-dependent (Peterson et al., 2011). At higher spatial resolution, biotic interactions become important, and corresponding biotic predictors will be needed (Pearson, Dawson, 2003; Austin, 2007). The spatial distributional pattern observed at different scales and grains may not be the same and may be controlled by different dominating factors (Guisan, Thuiller, 2005). Guisan and Thuiller (2005) note: "a gradual distribution observed over a large extent and at coarse resolution is likely to be controlled by climatic regulators, whereas patchy distribution observed over a smaller area and at fine resolution is more likely to result from a patchy distribution of resources, driven by micro-topographic variation or habitat fragmentation". In this study, we used ecologically relevant predictors which were selected based on the literature. Due to the presence of many mountains in Bulgaria, climatic conditions change rapidly over small distances. Therefore, climatic predictors such as

YRA are important to understand not only general patterns but also distribution at finer scales. The soil conditions, such as CF60, are another factor operating at a fine scale.

Despite best efforts, our approach to selecting explanatory environmental variables suffers from some degree of subjectivity. Alternative variables could be chosen, for instance, instead of GDD, mean annual temperatures could be used. The relevance of this variable was shown by Radkov (1970) who suggested that growth of Scots pine on sunny expositions is limited to areas with mean annual temperatures below 7°C. Also, instead of August precipitation other indicators of water availability and drought risk could be used. In particular, complex climate indices confronting thermal and humidity conditions are suitable. The aridity index of De Martonne is one such index and it was used to identify forest vulnerability zones in Bulgaria (Raev et al., 2015). Zimmermann et al., (2009) used a moisture index, which is the difference between precipitation and potential evapotranspiration. Another alternative, readily available in ClimateEU software, is the summer heat:moisture index (Marchi et al., 2020). Models usually exploit climatic data averaged over a multiannual period but extremes that occur during the observation period may provide important additional information for distribution modelling. Zimmermann et al., (2009) showed that measures representing climate extremes (i.e., interannual variability in climate parameters) explain and predict spatial patterns of 11 tree species in Switzerland. Soil characteristics such as soil richness and humidity are used in forestry practice to classify forest site types in terms of their quality and potential to support growing different tree species. This classification has been applied to identify the site types suitable for growing pine plantations (Kostov, 2014; Aleksandrov, Tonchev, 2021). It can, potentially, be used for SDM but maps of forest site types are not available outside the forest areas and thus could not be used with the current approach. Alternative would be to construct a similar site/soil map from other sources such as the SoilGrids 2.0 dataset and DEM-derived variables. For instance, field mapping of forest site types assesses humidity based on slope angle, exposition, and curvature, relief form, soil depth, degree of stoniness, and indicator species (Executive Forest Agency, 2011). At least part of these criteria can be obtained as digital datasets. Similarly, soil richness is assessed based on the bedrock, among other criteria, and this data can be obtained from geological maps.

SDMs, as applied in this study, are usually designated as correlative models to distinguish them from process-based (or mechanistic) distribution models (Dormann et al., 2011). Such SDM relies on the assumption that there is a (quasi-)equilibrium between the species and the physical characteristics of their environment (Guisan, Theurillat, 2000). These models are static, as they do not take into account the time dimension (Guisan, Theurillat, 2000); the modelling is for a specific moment, a snapshot of the relationship of the species with the environment (Guisan, Thuiller, 2005). If there is no equilibrium and the range of the species is in the process of adapting to the environment, then the snapshot will not be representa-

tive, and analyses conducted at different points in time will give different results. Another limitation of our model is that it does not include competition and other biotic interactions explicitly. Spatial datasets characterising such biotic factors are generally unavailable or available in resolution not suitable for modelling. A further complication, as discussed by Guisan and Theurillat (2000), is the fact that biotic interactions are intrinsically, but indirectly, integrated into the models. This is because the models are fed with empirical data characterising the observed species' distributions, which are, in reality, constrained by biotic interactions among other factors (Pearson, Dawson, 2003). The same is true also for the forestry data used in this study.

Future studies should experiment with other tree species for which data are available in the forest inventory databases. This may need adaptation of the methodology for species which are rarer or concentrated in a certain area. In particular, the size of the tiles used for cross-validation and the approach to deal with low prevalence may need careful reconsideration. In extreme cases, spatial block cross-validation may not be possible as most of the tiles would represent areas where the species is not present. Future studies may include different predictors or adopt variable selection procedures to improve prediction accuracy. Different tree species need an individual approach when selecting predictors. Other modelling methods, such as Generalised Additive Models and machine learning algorithms, should be explored too. Extending the methodology to other tree species would ultimately permit modelling species assemblages (biodiversity, composition) from individual species predictions.

Conclusions

An approach was demonstrated for tree species distribution modelling based on occurrence data from the stand-level forest inventory. It included filtering forest stands by area, geometric, and attributive criteria and then transforming each polygon to a representative point. The predictions of Scots pine distribution in Bulgaria using this derivative dataset had good accuracy (AUC=0.80, TSS=0.49). The model was relatively robust to spatial errors in the occurrence data as simulated by shifting the points randomly up to 1 km away from the original representative point. The ClimateEU software was particularly useful to derive climatic variables of high resolution (250 m), allowing integration with variables derived from DEM, and ultimately, significantly improving the spatial resolution in comparison with the common 1 km resolution of regional to continental models. Nevertheless, the approach has some drawbacks, in particular the assumption of homogeneity of forest stands, the unknown reliability of the inventory dataset, and the low prevalence (the absence points dominated significantly in number over the presence points). Future studies should focus on improving the accuracy and credibility of predictions by experimenting with different modelling methods and predictors.

Acknowledgements

We are grateful to the Executive Forest Agency and the directorates of the national parks "Rila", "Centralen Balkan", and "Pirin" for providing the forest inventory data. We thank the anonymous reviewer for the valuable comments and suggestions that helped improve the manuscript.

References

Aleksandrov N., Tonchev T. 2021. Ecosystem fit of Scots pine (*Pinus sylvestris* L.) plantations in south-western Bulgaria. Baltic Forestry 27 (2), 601. doi:10.46490/BF601.

Allouche O., Tsoar A., Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43, 1223–1232.

Armenteras D., Mulligan M. 2010. Modelling the potential distribution of tree species on a national scale in Colombia: Application to *Palicourea angustifolia* Kunth and *Palicourea guianensis* Aubl. Caldasia 32 (2), 355–380. https://revistas.unal.edu.co/index.php/cal/article/view/36229.

Austin M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling 200, 1–19. doi:10.1016/j.ecolmodel.2006.07.005.

Austin M.P., Nicholls A.O., Margules C.R. 1990. Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species. Ecological Monographs 60 (2), 161–177.

Baetens J.M., Verbist K., Cornelis W.M., Gabriels D., Soto G. 2009. On the influence of coarse fragments on soil water retention. Water Resources Research 45, W07408. doi:10.1029/2008WR007402.

Bencheva S., Doychev D. 2022. Winter desiccation of dwarf pine (*Pinus mugo* Turra) needles in the area of Belmeken Dam. Silva Balcanica 23 (1), 79–88. doi:10.3897/silvabalcanica.23.e90082.

Benito Garzón M., Blazek R., Neteler M., Sánchez de Dios R., Sainz Ollero H., Furlanello C. 2006. Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestris* L. in the Iberian Peninsula. Ecological Modelling 197, 383–393. doi:10.1016/j.ecolmodel.2006.03.015.

Bigler C., Bräker O. U., Bugmann H., Dobbertin M., Rigling A. 2006. Drought as an Inciting Mortality Factor in Scots Pine Stands of the Valais, Switzerland. Ecosystems 9, 330–343. doi:10.1007/s10021-005-0126-2.

Bohl C.L., Kass J.M., Anderson R.P. 2019. A new null model approach to quantify performance and significance for ecological niche models of species distributions. Journal of Biogeography 46, 1101–1111. doi:10.1111/jbi.13573.

Bolliger J., Kienast F., Bugmann H. 2000. Comparing models for tree distributions: concept, structures, and behaviour. Ecological Modelling 134, 89–102.

Bonannella C., Hengl T., Heisig J., Parente L., Wright M.N., Herold M., de Bruin S. 2022. Forest tree species distribution for Europe 2000–2020: mapping potential and realized distributions using spatiotemporal machine learning. PeerJ 10:e13728. doi:10.7717/peerj.13728.

Bondev I. 1991. The vegetation of Bulgaria: Map in 1:600 000 scale with explanatory text. Sofia University "St. Kliment Ohridski" press, Sofia, 184 pp. (In Bulgarian, English summary).

Bossche J. Van den, Jordahl K., Fleischmann M., McBride J., Wasserman J., Richards M., Badaracco A.G., Snow A.D., Tratner J., Gerard J., Ward B., Perry M., Farmer C., Hjelle G.A., Taves M., ter Hoeven E., Cochran M., Gillies R.S., Caria G., Culbertson L., Bell R., Bartos M., Eubank N., Flavin J., Rey S., Gardiner J., Bilogur A. 2023. geopandas/geopandas: v0.14.1 (v0.14.1), Zenodo. doi:10.5281/zenodo.2650956.

Brakensiek D.L., Rawls W.J. 1994. Soil containing rock fragments: effects on infiltration. Catena 23, (1/2), 99–110. doi:10.1016/0341-8162(94)90056-6.

Casalegno, S., G. Amatulli, A. Camia, A. Nelson, A. Pekkarinen. 2010. Vulnerability of Pinus cembra L. in the Alps and the Carpathian mountains under present and future climates. –Forest Ecology and Management, 259, 750–761.

Castro J., Zamora R., Hódar J.A., Gómez J.M. 2005. Ecology of seed germination of Pinus sylvestris L. at its southern, Mediterranean distribution range. Investigación Agraria: Sistemas y Recursos Forestales 14 (2), 143–152.

Chow T.L., Rees H.W., Monteith J.O., Toner P., Lavoie J. 2007. Effects of coarse fragment content on soil physical properties, soil erosion and potato production. Canadian Journal of Soil Science 87, 565–577.

Copernicus. 2024. Copernicus DEM - Global and European Digital Elevation Model. doi:10.5270/ESA-c5d3d65.

Dimitrov D., Zlatanov T., Raev I., Stoyanova N., Miteva S. 2011. Scots pine (*Pinus sylvestris* L.) response to climate changes and thinning activities: a tree-ring study from Southeast Rila mountain, Bulgaria. Silva Balcanica 12 (1), 63–70.

Dormann C.F., Schymanski S.J., Cabral J., Chuine I., Graham C., Hartig F., Kearney M., Morin X., Römermann C., Schröder B., Singer A. 2011. Correlation and process in species distribution models: bridging a dichotomy. Journal of Biogeography 39, 2119–2131. doi:10.1111/j.1365-2699.2011.02659.x.

Dunn P.K., Smyth G.K. 2018. Generalized Linear Models With Examples in R. Springer Texts in Statistics. Springer, New York, 562 pp.

Elith J., Leathwick J.R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics 40, 677–697. doi:10.1146/annurev.ecolsys.110308.120159.

Executive Forest Agency. 2011. Instructions for identifying and mapping forest site types and determining the composition of dendrocenoses. Sofia, BULPROFOR, 135 pp. (In Bulgarian).

Falk W., Mellert K.H. 2011. Species distribution models as a tool for forest management planning under climate change: risk evaluation of *Abies alba* in Bavaria. Journal of Vegetation Science 22, 621–634, doi:10.1111/j.1654-1103.2011.01294.x.

Franklin J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. Journal of Vegetation Science 9, 733–748.

Franklin J. 2009. Mapping species distribution: Spatial inference and prediction. Cambridge University Press. 320 pp.

GCCA. 2020. BGSTrans v4.6.2. Coordinate transformation application. Geodesy, Cartography, and Cadastre Agency.

Gogushev G., Zlatanov T. 2016. Structure and natural dynamics of *Pinus sylvestris* forests in Bulgaria. In: Panayotov M., Bebi P., Yurukov S. (Eds). Mountain coniferous forests in

Bulgaria – structure and natural dynamics. University of Forestry, Sofia, pp. 291–300. (In Bulgarian, English summary).

Graham C.H., Elith J., Hijmans R.J., Guisan A., Peterson A.T., Loiselle B.A. The Nceas Predicting Species Distributions Working Group. 2008. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45, 239–247. doi:10.1111/j.1365-2664.2007.01408.x.

Guisan A., Edwards Jr T.C., Hastie T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157, 89–100.

Guisan A., Theurillat J.-P. 2000. Equilibrium modeling of alpine plant distribution: how far can we go? Phytocoenologia 30 (3/4), 353–384. doi:10.1127/phyto/30/2000/353.

Guisan A., Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8, 993–1009. doi:10.1111/j.1461-0248.2005.00792.x.

Guisan A., Zimmermann N.E., Elith J., Graham C.H., Phillips S., Peterson A.T. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? Ecological Monographs 77 (4), 615–630.

Hijmans R.J. 2023. Terra: Spatial Data Analysis. R package version 1.7–23. https://CRAN.R-project.org/package=terra

Hijmans R.J. 2024. predicts: Spatial Prediction Tools. R package version 0.1–16. https://CRAN.R-project.org/package=predicts.

Hofierka J., Suri M. 2002. The solar radiation model for Open source GIS: implementation and applications. International GRASS users conference in Trento, Italy, September 2002.

Hosmer D.W, Lemeshow S. 2000. Applied Logistic Regression. Second Edition. Wiley, 375 pp.

Houston Durrant T., de Rigo D., Caudullo G. 2016. Pinus sylvestris in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J., de Rigo D., Caudullo G., Houston Durrant T., Mauri A. (Eds). European Atlas of Forest Tree Species. Publ. Off. EU, Luxembourg, pp. e016b94+.

Huntley B., Berry P.M., Cramer W., Mcdonald A.P. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. Journal of Biogeography 22(6), 967–1001.

Jiménez-Valverde A., Lobo J.M. 2007. Threshold criteria for conversion of probability of species presence to either–or presence–absence. Acta oecologica 31, 361–369.

Kostov G. 2014. An idea for classification of the Scots pine and Black pine plantations according to the level of ecosystem fit. Forest 9, 19–21. (In Bulgarian).

Kruskal W.H., Wallis W.A. 1952. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association 47 (260), 583–621. doi:10.1080/01621459.195.10483441.

Kullman L. 2007. Tree line population monitoring of Pinus sylvestris in the Swedish Scandes, 1973–2005: implications for tree line theory and climate change ecology. Journal of Ecology 95, 41–52. doi:10.1111/j.1365-2745.2006.01190.x.

Leathwick J.R., Austin M.P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. Ecology 82 (9), 2560–2573.

Marchi M., Ducci F. 2018. Some refinements on species distribution models using tree-level National Forest Inventories for supporting forest management and marginal forest population detection. iForest 11, 291–299. doi:10.3832/ifor2441-011.

Marchi M., Castellanos-Acuña D., Hamann A., Wang T., Ray D., Menzel A. 2020. ClimateEU, scale-free climate normals, historical time series, and future projections for Europe. Scientific Data 7, 428. doi:10.1038/s41597-020-00763-0.

Markoff I., Karamfilov K. 2016. Outlines of a national inventory of forests in Bulgaria. Forest Science (1/2), 83–101. (In Bulgarian, English summary).

Mauri A., Girardello M., Strona G., Beck P.S.A., Forzieri G., Caudullo G., Manca F., Cescatti A. 2022. EU-Trees4F, a dataset on the future distribution of European tree species. Scientific Data 9, 37. doi:10.1038/s41597-022-01128-5.

Meier E.S., Kienast F., Pearman P.B., Svenning J-C., Thuiller W., Araújo M.B., Guisan A., Zimmermann N.E. 2010. Biotic and abiotic variables show little redundancy in explaining tree species distributions. Ecography 33, 1038–1048. doi: 10.1111/j.1600-0587.2010.06229.x.

Nelder J.A., Wedderburn R.W.M. 1972. Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General) 135 (3), 370–384.

Nicholls A.O. 1989. How to make biological surveys go further with generalised linear models. Biological Conservation 50, 51–75.

Panayotov M., Yurukov S. 2016. Species composition and distribution of mountain coniferous forests in Bulgaria. In: Panayotov M., Bebi P., Yurukov S. (Eds). Mountain coniferous forests in Bulgaria – structure and natural dynamics. University of Forestry, Sofia, pp. 9–41. (In Bulgarian, English summary).

Panayotov M., Vassilev V., Tsvetkova N., Kovachev A. 2023. Ecological niches of main tree species in South-western Bulgaria and their modelling for expected future climatic conditions. In: Panayotov M., Tsvetanov N. (Eds). Climate changes and their effect on the forests in South-western Bulgaria. University of Forestry, Sofia, pp. 13–138. (In Bulgarian).

Pearson R.G., Dawson T.P. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? Global Ecology and Biogeography 12, 361–371.

Peterson A.T., Soberón J., Pearson R.G., Anderson R.P., Martínez-Meyer E., Nakamura M., Bastos Araújo M. 2011. Ecological Niches and Geographic Distributions. Princeton university press, 316 pp.

Poggio L., de Sousa L.M., Batjes N.H., Heuvelink G.B.M., Kempen B., Ribeiro E., Rossiter D. 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240. doi:10.5194/soil-7-217-2021.

Prasad A., Pedlar J., Peters M., McKenney D., Iverson L., Matthews S., Adams B. 2020. Combining US and Canadian forest inventories to assess habitat suitability and migration potential of 25 tree species under climate change. Diversity and Distributions 26, 1142–1159. doi:10.1111/ddi.13078.

Pshegusov R., Tembotova F., Chadaeva V., Sablirova Y., Mollaeva M., Akhomgotov A. 2022. Ecological niche modelling of the main forest-forming species in the Caucasus. Forest Ecosystems 9, 100019. doi:10.1016/j.fecs.2022.100019.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Radkov I.N. 1961. Ecological classification of forests in Rila Mountain. Sofia, Zemizdat, 202 pp. (In Bulgarian).

Radkov I.N. 1970. Ecological Foundations of Forestry. Sofia, Zemizdat, 323 pp. (In Bulgarian).

Raev I., Alexandrov V., Tinchev G. 2015. Assessment of drought related climate change impacts on forests in Bulgaria. Silva Balcanica 16 (1), 5–24.

Rehfeldt G.E., Tchebakova N.M., Parfenova Y.I., Wykoff W.R., Kuzmina N.A., Milyutin L.I. 2002. Intraspecific responses to climate in Pinus sylvestris. Global Change Biology 8, 912–929.

Sykes M.T., Prentice I.C,, Cramer W. 1996. A Bioclimatic Model for the Potential Distributions of North European Tree Species Under Present and Future Climates. Journal of Biogeography 23 (2), 203–233.

Theurillat J.-P., Felber F., Geissler P., Gobat J.-M., Fierz M., Fischlin A., Küpfer P., Schlüssel A., Velluti C., Zhao G.-F. 1998. Sensitivity of plant and soil ecosystems of the Alps to climate change. In: Cebon P., Dahinden U., Davies H.C., Imboden D., Jäger C.C. (Eds). Views from the Alps. Regional perspectives on climate change. MIT Press, London, pp. 225–308.

Varela S., Anderson R.P., García-Valdés R., Fernández-González F. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37. 001–008, doi:10.1111/j.1600-0587.2013.00441.x.

Vidal C., Alberdi I.A., Mateo L.H., Redmond J.J. (Eds). 2016. National Forest Inventories: Assessment of Wood Availability and Use. Springer Cham, 845 pp. doi:10.1007/978-3-319-44015-6.

Zimmermann N.E., Kienast F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. Journal of Vegetation Science 10, 469–482.

Zimmermann N.E., Edwards T.C., Moisen G.G, Frescino T.S., Blackard J.A. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. Journal of Applied Ecology 44, 1057–1067. doi: 10.1111/j.1365-2664.2007.01348.x.

Zimmermann N.E., Yoccoz N.G., Edwards T.C., Meier E.S., Thuiller W., Guisan A., D.R. Schmatz, Pearman P.B. 2009. Climatic extremes improve predictions of spatial patterns of tree species. Proceedings of the National Academy of Sciences of the United States of America, 106 (supplement_2), 19723–19728, doi:10.1073/pnas.0901643106.

Zimmermann N.E., Jandl R., Hanewinkel M., Kunstler G., Kölling C., Gasparini P., Breznikar A., Meier E.S., Normand S., Ulmer U., Gschwandtner T., Veit H., Naumann M., Falk W., Mellert K., Rizzo M., Skudnik M., Psomas A. 2013. Potential Future Ranges of Tree Species in the Alps. In: Management Strategies to Adapt Alpine Space Forests to Climate Change Risks. InTech, pp. 37–48/ doi:10.5772/56279.